

Lecture 2: Conditional Probability and Independence

EECS 126 (UC Berkeley)
Spring 2026 — January 22, 2026

In this lecture we introduce conditional probability, derive the multiplication (chain) rule, state the law of total probability and Bayes' rule, work through the classic false-positive medical testing example, and then develop the theory of independence and conditional independence.

1 Conditional Probability

Definition 2.1 (Conditional probability)

Let A, B be events with $\mathbb{P}(B) > 0$. The **conditional probability** of A given B is

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}. \quad (2.1)$$

Intuitively, $\mathbb{P}(A \mid B)$ is the probability that A occurs when we know that B has occurred. Conditioning on B restricts the sample space from Ω to B ; the event B becomes the new “ Ω ” for A .

In a uniform sample space, $\mathbb{P}(A) = |A|/|\Omega|$. Conditioning on B gives $\mathbb{P}(A \mid B) = |A \cap B|/|B|$, which is simply the fraction of outcomes in B that also lie in A .

Example 2.1 (Two dice – conditioning on the sum)

Roll two standard dice. Let $A = \{\text{1st die is 6}\}$, $B = \{\text{sum of two dice is 7}\}$, and $C = \{\text{sum of two dice is 11}\}$.

Computing $\mathbb{P}(A \mid B)$: The outcomes in B (sum = 7) are $\{(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1)\}$, so $\mathbb{P}(B) = 6/36$. The only outcome in $A \cap B$ is $(6, 1)$, so $\mathbb{P}(A \cap B) = 1/36$. Therefore

$$\mathbb{P}(A \mid B) = \frac{1/36}{6/36} = \frac{1}{6}. \quad (2.2)$$

Computing $\mathbb{P}(A \mid C)$: The outcomes in C (sum = 11) are $\{(5, 6), (6, 5)\}$, so $\mathbb{P}(C) = 2/36$. The only outcome in $A \cap C$ is $(6, 5)$, so

$$\mathbb{P}(A \mid C) = \frac{1/36}{2/36} = \frac{1}{2}. \quad (2.3)$$

2 Multiplication Rule (Chain Rule)

Theorem 2.1 (Chain rule of probability)

For any events $A_1, A_2, \dots, A_n$ with $\mathbb{P}(A_1 \cap A_2 \cap \dots \cap A_{n-1}) > 0$,

$$\mathbb{P}(A_1 \cap A_2 \cap \dots \cap A_n) = \mathbb{P}(A_1) \mathbb{P}(A_2 | A_1) \mathbb{P}(A_3 | A_1 \cap A_2) \dots \mathbb{P}(A_n | A_1 \cap \dots \cap A_{n-1}). \quad (2.4)$$

Proof. By repeated application of the definition of conditional probability. Write $\mathbb{P}(A_1 \cap A_2) = \mathbb{P}(A_1) \mathbb{P}(A_2 | A_1)$, then $\mathbb{P}(A_1 \cap A_2 \cap A_3) = \mathbb{P}(A_1 \cap A_2) \mathbb{P}(A_3 | A_1 \cap A_2)$, and continue inductively. Each factor is well-defined because $\mathbb{P}(A_1 \cap \dots \cap A_{k-1}) > 0$ implies $\mathbb{P}(A_1 \cap \dots \cap A_k) \leq \mathbb{P}(A_1 \cap \dots \cap A_{k-1})$, and we are given the intersection has positive probability up to $n - 1$ terms. $\square$

Birthday paradox via the chain rule. The chain rule gives an alternative derivation of the birthday bound. Let $A_i = \{\text{person } i \text{ doesn't share a birthday with persons } 1, \dots, i-1\}$. Then

$$\mathbb{P}(A_1 \cap \dots \cap A_n) = 1 \cdot \frac{K-1}{K} \cdot \frac{K-2}{K} \dots \frac{K-(n-1)}{K} = \prod_{j=0}^{n-1} \left(1 - \frac{j}{K}\right) \approx e^{-n(n-1)/(2K)}, \quad (2.5)$$

recovering the same bound from Lecture 1.

3 Law of Total Probability

Theorem 2.2 (Law of total probability)

Let $A_1, A_2, \dots, A_N$ be a **partition** of Ω (i.e. the A_i are pairwise disjoint — mutually exclusive — and their union is Ω — collectively exhaustive). Then for any event B ,

$$\mathbb{P}(B) = \sum_{i=1}^N \mathbb{P}(A_i) \mathbb{P}(B | A_i). \quad (2.6)$$

Proof. Since $\{A_i\}$ partitions Ω , we can decompose B as $B = \bigcup_{i=1}^N (A_i \cap B)$, and these pieces are disjoint. By countable additivity, $\mathbb{P}(B) = \sum_{i=1}^N \mathbb{P}(A_i \cap B) = \sum_{i=1}^N \mathbb{P}(A_i) \mathbb{P}(B | A_i)$, where the last step uses the definition of conditional probability. $\square$

4 Bayes' Rule

Theorem 2.3 (Bayes' rule)

Let $A_1, \dots, A_N$ partition Ω . Given an event B with $\mathbb{P}(B) > 0$,

$$\mathbb{P}(A_i | B) = \frac{\mathbb{P}(B | A_i) \mathbb{P}(A_i)}{\sum_{j=1}^N \mathbb{P}(B | A_j) \mathbb{P}(A_j)}. \quad (2.7)$$

The numerator uses the definition of conditional probability, and the denominator is $\mathbb{P}(B)$ expanded by the law of total probability.

Example 2.2 (False positive quiz – testing for a rare disease)

A medical test for a rare disease has the following properties:

- If a person *has* the disease, the test is positive with probability 0.95 (sensitivity).
- If a person *does not* have the disease, the test is negative with probability 0.95 (specificity), i.e. false positive rate = 0.05.
- The prior prevalence is $\mathbb{P}(A) = 0.001$ (1 in 1000 people have the disease).

A person tests positive. What is $\mathbb{P}(\text{disease} | \text{positive})$?

Define $A = \{\text{has disease}\}$, $B = \{\text{tests positive}\}$. By Bayes' rule:

$$\mathbb{P}(A | B) = \frac{\mathbb{P}(B | A) \mathbb{P}(A)}{\mathbb{P}(B | A) \mathbb{P}(A) + \mathbb{P}(B | A^c) \mathbb{P}(A^c)} = \frac{(0.95)(0.001)}{(0.95)(0.001) + (0.05)(0.999)} \approx 0.0187. \quad (2.8)$$

Despite the test being 95% accurate, a positive result only indicates a $\approx 1.87\%$ chance of actually having the disease. This is because the disease is so rare that the false positives from healthy people vastly outnumber the true positives.

Sensitivity to the prior. If the prior were $\mathbb{P}(A) = 0.01$ instead ($10\times$ more common), then $\mathbb{P}(A | B) \approx 16\%$ — still low, but dramatically higher. The base rate (prior) has an enormous effect on the posterior.

5 Independence

Definition 2.2 (Independence of two events)

Two events A and B are **independent** if the occurrence of one provides no information about the occurrence of the other. Formally:

$$\mathbb{P}(A | B) = \mathbb{P}(A), \quad \text{or equivalently,} \quad \mathbb{P}(A \cap B) = \mathbb{P}(A) \mathbb{P}(B). \quad (2.9)$$

The second form says the joint probability equals the product of the marginals.

Key remarks on independence:

1. **Disjoint $\neq$ independent.** A, B disjoint means $\mathbb{P}(A \cap B) = 0$. For this to equal $\mathbb{P}(A)\mathbb{P}(B)$, we need $\mathbb{P}(A) = 0$ or $\mathbb{P}(B) = 0$. So disjoint events with positive probability are *dependent*, not independent.
2. **Base outcomes are always dependent.** Individual outcomes $\{\omega\}$ in any finite Ω are

always dependent (knowing one occurred rules out the others).

Example 2.3 (Dice: independence depends on the conditioning event)

Roll two fair dice. Let $A = \{\text{1st die is 6}\}$, $B = \{\text{sum is 7}\}$, $C = \{\text{sum is 11}\}$.

Are A and B independent? $\mathbb{P}(A | B) = 1/6 = \mathbb{P}(A)$, so **yes**, $A \perp B$.

Are A and C independent? $\mathbb{P}(A | C) = 1/2 \neq 1/6 = \mathbb{P}(A)$, so **no**, A and C are dependent. Intuitively, knowing the sum is 11 makes it more likely that the first die was a 6 (since (6, 5) is one of only two outcomes).

6 Conditional Independence

Definition 2.3 (Conditional independence)

Events A and B are **conditionally independent** given C if

$$\mathbb{P}(A \cap B | C) = \mathbb{P}(A | C) \mathbb{P}(B | C). \quad (2.10)$$

Independence and conditional independence are distinct concepts:

1. Dependent events can become conditionally independent.
2. Independent events can become conditionally dependent.

Example 2.4 (Two indistinguishable coins)

You have two coins: one is double-headed, the other is double-tailed. You pick one at random and flip it twice. Let $H_i = \{\text{flip } i \text{ is heads}\}$, $i = 1, 2$.

Q1: Are H_1, H_2 (unconditionally) independent? No. $\mathbb{P}(H_1 | H_2) = 1 \neq 1/2 = \mathbb{P}(H_1)$. If the second flip is heads, you certainly picked the double-headed coin, so the first flip is heads with certainty.

Q2: Conditioning on the coin, are they independent? Let $A = \{\text{picked the double-headed coin}\}$. Then $\mathbb{P}(H_1 \cap H_2 | A) = \mathbb{P}(H_1 | A) \mathbb{P}(H_2 | H_1 \cap A) = 1 \cdot 1 = 1$, and $\mathbb{P}(H_1 | A) \mathbb{P}(H_2 | A) = 1 \cdot 1 = 1$. These are equal, so $H_1 \perp H_2 | A$. Once we know which coin was picked, each flip is determined and independent.

Example 2.5 (Fair coin – independent but conditionally dependent)

Toss a fair coin twice. Let $H_1 = \{\text{1st toss is H}\}$, $H_2 = \{\text{2nd toss is H}\}$, and $D = \{\text{1st toss} \neq \text{2nd toss}\}$.

Q1: $H_1 \perp H_2$? Yes, by independence of fair coin flips.

Q2: $H_1 \perp H_2 | D$? No. $\mathbb{P}(H_1 | D) \mathbb{P}(H_2 | D) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$. But $\mathbb{P}(H_1 \cap H_2 | D) = 0$, since both flips being heads contradicts D (the flips must differ). Since $0 \neq 1/4$, the events are conditionally dependent given D .

7 Joint Independence vs. Pairwise Independence

Definition 2.4 (Independence of a collection)

Events $\{A_i\}_{i \in S}$ are **jointly (mutually) independent** if for every subset $T \subseteq S$,

$$\mathbb{P}\left(\bigcap_{i \in T} A_i\right) = \prod_{i \in T} \mathbb{P}(A_i). \quad (2.11)$$

Pairwise independence only requires this for all pairs $|T| = 2$.

Pairwise independence $\nRightarrow$ joint independence. Consider the fair coin example: H_1, H_2, D are pairwise independent (check all three pairs), but they are *not* jointly independent because $\mathbb{P}(H_1 \cap H_2 \cap D) = 0 \neq \mathbb{P}(H_1)\mathbb{P}(H_2)\mathbb{P}(D) = 1/8$.

8 Renewal: Rolling a 6 Before a 7

Example 2.6 (Roll a sum of 6 before a sum of 7)

Roll two fair dice independently and repeatedly. Let $E = \{\text{a sum of 6 appears before a sum of 7}\}$.

The outcomes that sum to 6 are $\{(1, 5), (2, 4), (3, 3), (4, 2), (5, 1)\}$, so $\mathbb{P}(S) = 5/36$ where $S = \{\text{1st roll sums to 6}\}$. Similarly, $\mathbb{P}(T) = 6/36$ where $T = \{\text{1st roll sums to 7}\}$.

By the law of total probability, conditioning on the first roll:

$$\mathbb{P}(E) = \mathbb{P}(E | S) \mathbb{P}(S) + \mathbb{P}(E | T) \mathbb{P}(T) + \mathbb{P}(E | (S \cup T)^c) \mathbb{P}((S \cup T)^c). \quad (2.12)$$

We have $\mathbb{P}(E | S) = 1$, $\mathbb{P}(E | T) = 0$, and $\mathbb{P}(E | (S \cup T)^c) = \mathbb{P}(E)$ by the renewal/memoryless property (each roll is independent, so if neither 6 nor 7 appears, we restart). Thus

$$\mathbb{P}(E) = 1 \cdot \frac{5}{36} + 0 \cdot \frac{6}{36} + \mathbb{P}(E) \cdot \frac{25}{36}. \quad (2.13)$$

Solving: $\mathbb{P}(E)(1 - 25/36) = 5/36$, so $\mathbb{P}(E)(11/36) = 5/36$, giving

$$\mathbb{P}(E) = \frac{5}{11}. \quad (2.14)$$